

Sample: SRR23410316

FASTQ Quality

Filename	SRR23410316_R1.fastq.gz	SRR23410316_R2.fastq.gz
File Size	74.4 MB	74.1 MB
Q30 Passing	94.77%	93.69%
Mean Read Score	33.4	28.7
Average Read Length	149.1	149.1

Read Mapping against SRR23410316

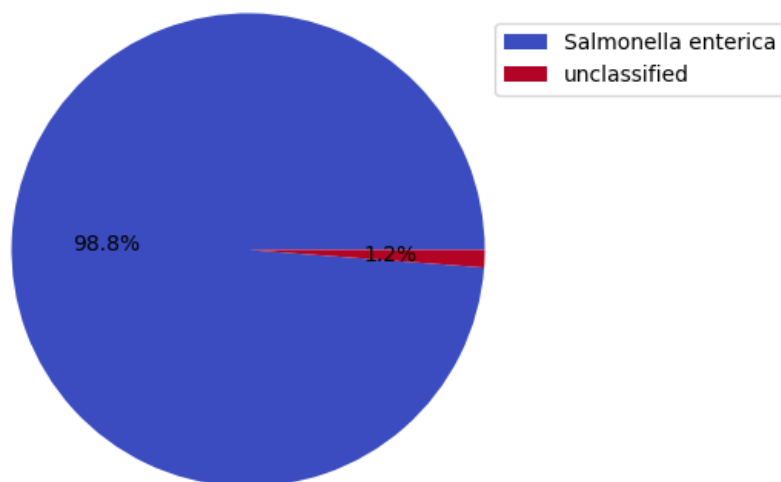
Mapped Paired Reads	Mapped Single Reads	Unmapped Reads	Unmapped Percent	Unmapped Assembled Contigs	
660,353	965	1,685	0.3%	assembly not done	
Duplicate Paired Reads	Duplicate Single Reads	Duplicate Percent of Mapped Reads			
13,523	584	2.1%			
BAM File	Reference Length	Genome with Coverage	Average Depth	No Coverage Bases	Quality SNPs
SRR23410316_nodup.bam	4,893,754	99.94%	39.3X	2,699	233

Assembly

Contig count	Contig length counts < 301-999bp >	Longest contig	Total length	N50	FASTQ calculated mean coverage
105	56 18 31	907,691	4,893,754	380,928	40.3X

FASTQ Identifications

FASTQ Read Identification



Identified using: Kraken and Bracken

Genome Size Comparisons

Salmonella enterica	expected: 5,050,000 nt
SRR23410316	assembled: 4,893,754 nt

Sourmash Sequence Similarity

Similarity	ID
97.1%	MXLR01000010.1 Salmonella enterica subsp. enterica serovar Newport strain INSP 15 NODE_10_length_140122_cov_5.59513, whole genome shotgun sequence
96.9%	MXUZ01000100.1 Salmonella enterica strain BCW_3053 NODE_100_length_249_cov_0.401235, whole genome shotgun sequence
95.2%	CP007559.1 Salmonella enterica subsp. enterica serovar Newport str. CDC 2010K-2159, complete genome
94.2%	AHTR01000001.1 Salmonella enterica subsp. enterica serovar Newport str. CVM 19567 SEEN567_10, whole genome shotgun sequence
94.1%	MYAV01000010.1 Salmonella enterica subsp. enterica serovar Newport strain BCW_2742 NODE_10_length_203084_cov_2.87447, whole genome shotgun sequence
94.0%	AHTQ01000001.1 Salmonella enterica subsp. enterica serovar Newport str. CVM 19449 SEEN449_10, whole genome shotgun sequence
93.9%	AHTP01000001.1 Salmonella enterica subsp. enterica serovar Newport str. CVM 19447 SEEN447_10, whole genome shotgun sequence
93.8%	AOGH01000001.1 Salmonella enterica subsp. enterica serovar Newport str. Henan_3 contig001, whole genome shotgun sequence
89.0%	MYSQ01000010.1 Salmonella enterica subsp. enterica serovar Newport strain BCW_4368 NODE_10_length_200426_cov_4.64018, whole genome shotgun sequence
88.5%	MYSQ01000010.1 Salmonella enterica subsp. enterica serovar Newport strain BCW_4378 NODE_10_length_175619_cov_2.92217, whole genome shotgun sequence

MLST 2.19.0

Schema-Sequence type: **senterica-132**

MLST Detail: aroC(2), dnaN(57), hemD(15), hisD(14), purE(15), sucA(20), thrA(12)

Results provided by: MLST

AMRFinderPlus - version 3.11.2, database 2021-03-01.1

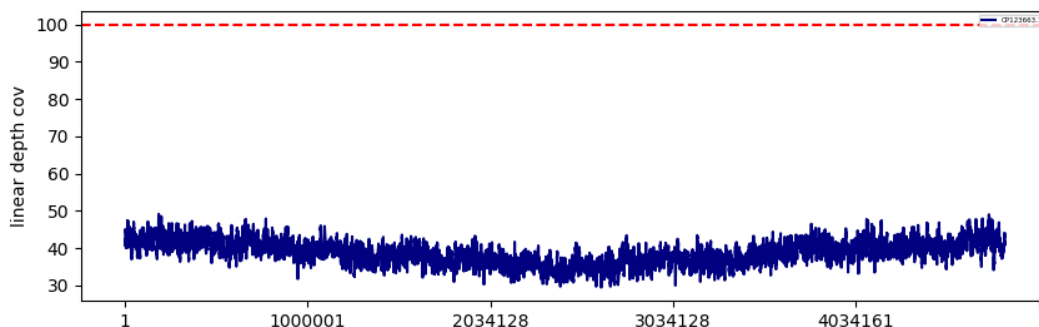
AMRFinderPlus Organism Option: Salmonella

Contig ID	Element	Gene	Description	% Cov	% Similar
NODE_18_length_23727_cov_21.218816	AMR	mph(A)	Mph(A) family macrolide 2'-phosphotransferase	100.0%	99.67%
NODE_18_length_23727_cov_21.218816	AMR	sul1	sulfonamide-resistant dihydropteroate synthase Sul1	100.0%	100.0%
NODE_18_length_23727_cov_21.218816	BIOCIDE	qacEdelta1	quaternary ammonium compound efflux SMR transporter QacE delta 1	100.0%	100.0%
NODE_18_length_23727_cov_21.218816	AMR	aadA2	ANT(3'')-Ia family aminoglycoside nucleotidyltransferase AadA2	100.0%	100.0%
NODE_18_length_23727_cov_21.218816	AMR	blaCARB-2	PSE family carbenicillin-hydrolyzing class A beta-lactamase CARB-2	100.0%	100.0%
NODE_18_length_23727_cov_21.218816	AMR	qnrA1	quinolone resistance pentapeptide repeat protein QnrA1	100.0%	100.0%
NODE_18_length_23727_cov_21.218816	AMR	sul1	sulfonamide-resistant dihydropteroate synthase Sul1	100.0%	100.0%
NODE_18_length_23727_cov_21.218816	BIOCIDE	qacEdelta1	quaternary ammonium compound efflux SMR transporter QacE delta 1	100.0%	100.0%
NODE_18_length_23727_cov_21.218816	AMR	dfrA1	trimethoprim-resistant dihydrofolate reductase DfrA1	100.0%	100.0%
NODE_18_length_23727_cov_21.218816	AMR	tet(A)	tetracycline efflux MFS transporter Tet(A)	100.0%	100.0%
NODE_18_length_23727_cov_21.218816	AMR	floR	chloramphenicol/florfenicol efflux MFS transporter FloR	100.0%	99.75%
NODE_1_length_907691_cov_19.701392	VIRULENCE	iroC	salmonchelin/enterobactin export ABC transporter IroC	98.85%	79.92%
NODE_1_length_907691_cov_19.701392	VIRULENCE	iroB	salmonchelin biosynthesis C-glycosyltransferase IroB	100.0%	85.98%
NODE_21_length_21041_cov_17.631750	VIRULENCE	sodC1	superoxide dismutase [Cu-Zn] SodC1	100.0%	100.0%
NODE_3_length_524566_cov_18.983327	METAL	golS	Au(I) sensor transcriptional regulator GolS	100.0%	100.0%
NODE_3_length_524566_cov_18.983327	METAL	golT	gold/copper-translocating P-type ATPase GolT	100.0%	99.61%
NODE_3_length_524566_cov_18.983327	AMR	mdsA	multidrug efflux RND transporter periplasmic adaptor subunit MdsA	100.0%	98.28%
NODE_3_length_524566_cov_18.983327	AMR	mdsB	multidrug efflux RND transporter permease subunit MdsB	100.0%	99.72%
NODE_5_length_380928_cov_18.153469	VIRULENCE	sinH	intimin-like inverse autotransporter SinH	100.0%	99.86%

Results provided by: AMRFinderPlus

Coverage Graph

SRR23410316 reads aligned against CP123663



ID	Average Depth	Genome Coverage	Genome Length	Ambiguous SNPs
CP123663.1	38.2X	98.7%	4,880,046	29

ID Detail

CP123663.1 *Salmonella enterica* subsp. *enterica* serovar Newport strain FSIS22130757 chromosome, compl

SeqSero2 Detail

Predicted serotype(s)	Newport
Predicted antigenic profile	8:e,h:1,2
Predicted subspecies	<i>Salmonella enterica</i> subspecies <i>enterica</i> (subspecies I)

Note:

Header	Antigen	Blast Score	Identity
NODE_2_length_1446_cov_34.153397	O-8_wzy	2211.56	99.92%
NODE_1_length_1748_cov_35.073609	fliB 1,2	2760.01	99.41%
NODE_3_length_1224_cov_26.126417	fliC e,h	2050.9	74.09%

Results provided by: SeqSero2

Sistr

Serovar

Sistr ran but not completed

PlasmidFinder Detail

Plasmid	Description	Length	Contig ID	HSP Length	% Identity	% Coverage
IncX1	EU370913.1 <i>Escherichia coli</i> plasmid pOLA52, comple	51,602	NODE_15_length_35422_cov_31.569105	374	94.92%	100.0%
IncR	DQ449578.1 <i>Klebsiella pneumoniae</i> strain NK245 plas	98,264	NODE_22_length_15973_cov_21.623113	251	100.0%	100.0%

Results provided by: PlasmidFinder

BLAST nt - Assembly Identification

Nucleotide Representa- tion	Description
1,553,865	CP123663 <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Newport strain FSIS22130757 chromosome, complet
907,691	CP119463 <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Newport strain FSIS32104969 chromosome, complet
580,535	CP123669 <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Newport strain FSIS12142912 chromosome, complet
568,874	CP123671 <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Newport strain FSIS12105828 chromosome, complet
524,566	CP119501 <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Newport strain FSIS12034723 chromosome, complet
380,577	CP123686 <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Newport strain FSIS11814458 chromosome, complet
75,579	CP119514 <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Newport strain FSIS31903059 chromosome, complet
7,101	CP025232 <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Newport str. USDA-ARS-USMARC-1925 chromosome, c
6,532	CP041208 <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Newport strain SAP18-8729 chromosome, complete
4,752	CP043753 <i>Escherichia coli</i> strain CVM N56639 plasmid pN56639, complete sequence
3,130	CP030175 <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Milwaukee str. SA19950795 chromosome, complete
2,211	CP055130 <i>Salmonella enterica</i> strain FDAARGOS_711 chromosome
1,993	CP054901 <i>Salmonella enterica</i> strain FDAARGOS_718 chromosome
1,514	AP020330 <i>Salmonella enterica</i> SEHaa3795 DNA, complete genome
1,160	CP053702 <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhi strain CMCST_CEPR_1 chromosome, complet
1,104	CP053402 <i>Salmonella enterica</i> strain 2016K-0213 chromosome, complete genome
1,080	LR134140 <i>Salmonella enterica</i> subsp. <i>enterica</i> strain NCTC129 genome assembly, chromosome: 1
960	CP036168 <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhimurium strain sg_wt7 chromosome, complete
895	CP055292 <i>Shigella sonnei</i> strain SE6-1 chromosome, complete genome
822	CP054380 <i>Escherichia coli</i> strain SCU-175 plasmid pSCU-175-1, complete sequence
658	CP051273 <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Worthington strain OLF-FSR1_WB_Partridge_SW-
630	CP054897 <i>Salmonella enterica</i> strain FDAARGOS_717 chromosome
535	CP053332 <i>Salmonella enterica</i> subsp. <i>houtenae</i> serovar 44:z4,z32:- strain 2009K1701 chromosome, comple
528	CP052181 <i>Klebsiella pneumoniae</i> strain F16KP0037 chromosome, complete genome
371	CP082478 <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Newport strain FSIS1607290 chromosome, complete
290	LR134147 <i>Salmonella enterica</i> subsp. <i>enterica</i> strain NCTC7831 genome assembly, chromosome: 1
280	MT219826 <i>Escherichia coli</i> strain RW8-1 plasmid pRW8-1_122k_tetX, complete sequence
256	CP045955 <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Enteritidis strain AUSMDU00010528 chromosome, c
251	CP050716 <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Enteritidis strain SE95 chromosome, complete ge
215	CP051270 <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Worthington strain OLF-FSR1_WB_Quail_SW-70 c
186	AP020332 <i>Salmonella enterica</i> SESen3709 DNA, complete genome
165	CP053585 <i>Salmonella enterica</i> strain 2011K-1440 chromosome, complete genome
165	CP049987 <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Saintpaul strain CVM N16S133 plasmid pN16S133,
155	CP053406 <i>Salmonella enterica</i> strain 68-4603 chromosome, complete genome
150	CP053214 <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Kedougou strain Sal162 chromosome
122	CP052366 <i>Klebsiella pneumoniae</i> strain D16KP0109 plasmid pD16KP0109-1, complete sequence
104	CP037917 <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Kentucky strain K13SK002 chromosome
99	CP054715 <i>Salmonella enterica</i> strain 85-0120 chromosome, complete genome
91	CP082547 <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Newport strain CVM N18S1655 chromosome, complet

Results provided by: BLAST nt database

Report Description

Input to create this report was either FASTA, or single/paired FASTQ/s.

If built from FASTA, contigs have been identified using Kraken2/Bracken; and assembly metrics, expected genome size compared against total assembly length, MLST, PlasmidFinder, AMRFinderPlus and BLAST results have been reported. In addition to these tests, if report was built from FASTQ/s, sequence technology, FASTQ metrics and coverage graph against the top BLAST hit was also reported. Finally, if Kraken2/Bracken top finding was Salmonella SeqSero2 and CRISPR SeroSeq additionally reported. See table below for summary of tests reported.

Expected results are approximately based on a 4,500,000 base bacterial genome.

When tests are run and there are no significant findings the test's banner and a "no results" message will still be included in this report. If this is not seen, yet test is expected, the report is invalid.

	FASTA	FASTQ	FASTQ Identified as Salmonella
Sequence Technology		X	X
FASTQ metrics		X	X
Kraken2/Bracken	X	X	X
Assembly metrics	X	X	X
Expected Genome Size	X	X	X
MLST	X	X	X
PlasmidFinder	X	X	X
AMRFinderPlus	X	X	X
Coverage Graph		X	X
BLAST	X	X	X
SeqSero2			X
CRISPR SeroSeq			X

Table 1: Test Summary

Sequence Technology

Single FASTQ is assumed to have been generated from Ion S5 System using S5 chemistry. Paired reads are assumed to have been generated from Illumina MiSeq System. Read length is reported based on average read length and rounded up to 50th number.

FASTQ metrics

Filename: Input files

File Size: Compressed input file size (.gz). *Expected: >150MB*

Q30 Passing: Percentage of reads with an average Phred score above 30. *Expected: >70%*

Mean Read Score: Average Phred score of all base calls. *Expected: >30*

Kraken2/Bracken

Bracken computes specie abundance using Kraken2 taxonomy labels. Fraction of reads are reported only when greater than 1% of total reads.

Kraken2 database was built from the following:

Libraries:

archaea: RefSeq complete archaeal genomes
 bacteria: RefSeq complete bacterial genomes
 plasmid: RefSeq plasmid nucleotide/protein sequences
 viral: RefSeq complete viral genomes
 human: GRCh38 human genome
 fungi: RefSeq complete fungal genomes
 plant: RefSeq complete plant genomes
 protozoa: RefSeq complete protozoan genomes
 UniVec Core: A subset of UniVec chosen to minimize false positive hits to the vector database

Custom genomes:

Bos taurus (cow)
 Sus scrofa (pig)
 gallus gallus (chicken)
 Odocoileus virginianus (white-tailed deer)
 Panthera tigris altaica (tiger)
 Felis catus (house cat)
 Canis lupus familiaris (dog)
 Loxodonta africana (elephant)
 Tursiops truncatus (bottlenose dolphin)
 Anas platyrhynchos (mallard)
 Cyprinus carpio (common carp)
 Nanorana parkeri (frog)
 Aedes albopictus (mosquito)

Unclassified reads were either poor quality or no perfect sample k-mers (k=35) mapped to read sequence from the Kraken2 database build.

Assembly metrics

Reads are assembled using SPAdes when FASTQs are used to generate report. When a FASTA is used SPAdes is assumed assembler.

Metrics can vary greatly. Numbers outside expected values should not always be interpreted as a poor sample, but should be considered against other test results.

Scaffolds: Total contig count of assembly. *Expected: <300*

Total Length: Total length of all sequences. *Expected: 4,500,000*

Longest Scaffold: Longest assembled sequence. *Expected: >200,000*

Scaffolds >1K nt: Sequences greater than 1,000 nucleotides in length. *Expected: >200*

N50: Half of the total assembled length is made from sequence greater than this value. *Expected: >100,000*

Mean Read Depth: Estimated average read depth across total assembly length. *Expected: >50X*

Genome Size Comparisons

Expected genome size is found using top Bracken identification. The genome size is estimated by providing the species name to NCBI which returns the expected size. If NCBI does not return a genome size for the top Bracken identification the Bracken list is iterated from highest to lowest read identification until genome size is found. If an expected genome size is still not found a value of 5,000,000 is used.

The comparison between sample assembly size and expected genome size is very important when determining contamination and genome completeness.

MLST

Assembly file scanned for PubMLST typing schemes using MLST tool.

Schema-Sequence type: Traditional PubMLST typing scheme found

MLST Detail: Gene and allele IDs, see link for explanations for "n", "~", "?", "-" and "n, m" notation.

PlasmidFinder

PlasmidFinder tool used to target plasmids.

Description: NCBI accession number and identification

Length: Sequence length of described accession

Contig ID: Assembly identification of found plasmid

HSP Length: High score pairs length of BLAST database sequence identifying plasmid

% Identity: Percentage of HSP bases similar to matched findings

% Coverage: Percentage of HSP bases aligning to matched findings

AMRFinderPlus

AMRFinderPlus finds acquired antimicrobial resistance genes in assembled nucleotide sequences and resistance-associated point mutations in select taxa. Gene classes found include AMR, point mutations, virulence factors, biocide, heat, acid, and metal resistance genes.

This report includes providing an organism option to AMRFinderPlus if Bracken identification fits select options. When an organism option is provided point mutations are screened and some common findings are suppressed.

AMRFinderPlus Organism Option: Option used to get organism-specific results

Contig ID: Assembly identification of found element

Element: Classification of the AMRFinderPlus gene

Gene: AMR gene

Name: Gene detail

% Cov: Percentage of protein aligning to matched findings

% Similar: Percentage of protein similar to matched findings

Coverage Graph

Reads aligned using BWA to the top BLAST accession hit. When greater than 100X depth-of-coverage log read depth is graphed. If less than 100X a linear read depth is graphed.

ID: NCBI accession aligned against

Average Depth: Average depth-of-coverage across sequence aligned against

Genome Length: Length of genome aligned against

Ambiguous SNPs: Ambiguous SNP count. A high count can indicate poor WGS (i.e. low coverage), a multiple-strain sample or contamination with a closely related strain.

ID Detail: Accession detail

BLAST

Assembly BLAST against nt database. Totaled nucleotide representation identified as top hit.

Nucleotide Representation: Summed contig lengths based on same accession top hit identifications.

Description: NCBI accession number and identification

SeqSero2

SeqSero2 targets the O (wzx and wzy genes) and H (fliC and fljB genes) antigens to determine serotype. When more than one O antigen, or two or more H antigens are found it can indicate the presence of contamination.

Predicted serotype(s): Predicted serotype based on O and H antigens

Predicted antigenic profiles: Shorthand profiles

Predicted subspecies: Predicted subspecies if available

Note: Special notes from SeqSero when provided

Bottom: Contig detail of SeqSero findings